

Chi-Square:

- * for categorical data
- * yesterday... one sample of data
- * today... two samples of data
- * always testing whether observed (sample) fits expected (population)

9.1: Chi Square Test for Association

Review: 2-way tables

Comparing gender of AP Stat students to their college location choices

	<i>gender</i>		
	Female	Male	total
In-State	17	21	38
Out of State	9	18	27
Total	26	39	65 65

college location

$$r = 2$$

$$c = 2$$

$$\text{cells} = 4$$

$$P(F \cap I_n)$$

$$P(F | I_n)$$

Question: Is there an association between the row and column variables?

- Compare... observed cell counts to the expected cell counts
- If... no association, $obs \approx expected$
- Use... χ^2 test for association (independent)

Question: How do we find expected values?

$$\text{Expected Cell Count} = \frac{\text{row total} \times \text{column total}}{n}$$

Find the expected cell counts for the table of college choices versus gender:

	Female	Male	total
In-State	15.2	22.8	38
Out of State	10.8	16.2	27
Total	26	39	65

Chi-Square Test for Association

- Want to test... if there is an association btw. row & column variables

Hypotheses:

$\textcircled{=}$ • H_0 : there is no association btw. row & column variables. (indep)

$\textcircled{\neq}$ • H_a : there is an association btw. row & column. (dependent)

o No association = independent

o Fill in... row & column var. names

Test Statistic:

- Same:

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = \frac{(\quad)^2}{\quad} + \frac{(\quad)^2}{\quad} + \dots =$$

- Compare... obs. + exp. cells

P-Value:

$$P(\chi^2 > \underline{\hspace{2cm}} \mid df = \quad) =$$

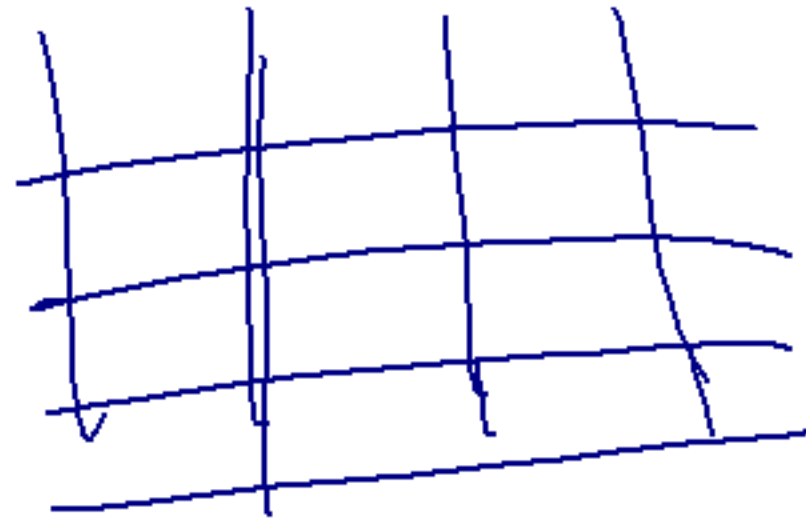
- Still... χ^2 distrib.
- $df = (r-1)(c-1)$
- Calculator: $\chi^2_{cdf}(LB, UB, df)$

Conclusion:

- Same...
 - reject/fail
 - recopy H_0/H_a

On Calculator:

- χ^2 test
- matrix
-



Assumptions:

- 2 indep. SRS
- all exp. counts ≥ 5 ✓

1. The manager of an assembly process wants to determine whether the number of defective articles manufactured depends on the day of the week the articles are produced. Using the data below, is there sufficient evidence to determine if the number of defective articles is independent of the day of the week?

Day	Mon.	Tue.	Wed.	Thur.	Fri.
Nondef.	85	90	95	95	90
Defective	15	10	5	5	10

H_0 : there is no association btw. defective articles & days of the week.

H_a : there is an association btw. defective articles & days of the week.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = \frac{(85 - 91)^2}{91} + \frac{(90 - 91)^2}{91} + \dots = 8.547$$

$$P(\chi^2 > 8.547 | df = 4) = 0.07$$

- We fail to reject H_0 b/c p-value $> \alpha = 0.05$.
- We have suff. evid. that... (recopy H_0).

- ② H_0 : the type of irritation and age are independent
 H_a : the type of irritation and age are dependent.

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} = \frac{(440 - 432.9)^2}{432.9} + \frac{(567 - 585.97)^2}{585.97} + \dots = 13.619$$

$$P(\chi^2 > 13.619 \mid df = 6) = 0.0342$$

- We reject H_0 b/c p-value $< \alpha = 0.05$.
- We have suff. evid. that the type of irritation & age are dependent.

STATE

CHECK

- ① 2 indep SRS ① assumed
- ② all exp. counts ≥ 5 ② ✓

$$\textcircled{3} \chi^2 = 337.662$$

$$df = 20$$

$$P\text{-value} = 1.5434 \times 10^{-59}$$

For # 3 and 4:

- don't check assumptions**
- you can abbreviate the conclusions**
 - reject or fail to reject**
 - write re-copy H_0 or H_a**